OIPE

#5

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/637,780B

DATE: 11/20/2001
TIME: 09:18:49

Input Set : N:\EBONY'S\ES.txt
Output Set: N:\CRF3\11202001\I637780B.raw

ENTERED

```
 3 <110> APPLICANT: ALEXANDROV, Nickolai
 4        BROVER, Vyacheslav
 6 <120> TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
Polypeptides Encoded
 7        Thereby
 9 <130> FILE REFERENCE: 2750-1096P
11 <140> CURRENT APPLICATION NUMBER: 09/637,780B
12 <141> CURRENT FILING DATE: 2000-08-11
14 <160> NUMBER OF SEQ ID NOS: 1401
16 <210> SEQ ID NO: 1
17 <211> LENGTH: 1458
18 <212> TYPE: DNA
19 <213> ORGANISM: Arabidopsis thaliana
21 <220> FEATURE:
22 <221> NAME/KEY: misc_feature
23 <222> LOCATION: 1..1458
24 <223> OTHER INFORMATION: any n = a, g, c, t, unknown, or other
26 <220> FEATURE:
27 <221> NAME/KEY: misc_feature
28 <222> LOCATION: 1..1458
29 <223> OTHER INFORMATION: Ceres Seq. ID 1481332
31 <400> SEQUENCE: 1
33 atccgtttcg ccatttttgt ttctcagtga tctctgaaat ggtctcttct cttttttttgg      60
34 tcgaatccaa tctcaattat gttgttatct ttcttccatc aatgggtaat caaaacatag     120
35 aattgatgcg gtaagactat aaaggtttag tctttaacca ttgtagattc ctctgtctct     180
36 tgtgtatttg attgatctgt taatggataa ccaaaaaggt gctctctttc ccgatgaggt     240
37 tattctccag attcttgcta gattaccgt taaatctctc ttcaggttca aatccgtttg     300
38 caaatcatgg tacagattac cttctgacaa atatttcact tccttgttca atcaactctc     360
39 tgtaaaagag caattgcttg tggctcaagt atcagattct tctagtttga tctgtgttga     420
40 taatctgaga ggtgtttctg agttatcatt ggattttgtt agagataggg tgaggattag     480
41 ggtttcttct aatggtttgt tgtgttgttc aagcattcct gaaaagggtg tttactatgt     540
42 ttgtaatccg tcgactagag agtacaggaa attgcctaag agtcgagaaa gacccgttac     600
43 tcggttttat cctgacggtg aggctacact tgttggtttg gcttgtgatt tgagtaggaa     660
44 caagtttaat gtggtgttgg ctggttacca taggtctttt ggtcagagac ctgatgggag     720
45 tttcatttgc ttggtgtttg attctgagag taacaaatgg aggaagtttg tttcggtgtt     780
46 agaagaatgt agtttcacac acatgagtaa gaaccaagtg gtgtttgtta atgggatgct     840
47 tcattggttg atgagtggtt tgtgttatat acttgcactt gatgttgaac atgatgtgtg     900
48 gagaaagatt tctttgcctg atgagattaa aatcgggaat ggtggtggta atcgggttta     960
49 tctcttggaa tccgatgggt ttttgtcggt gattcagtta tcagatgtat ggatgaagat    1020
50 ttggaagatg agtgagtatg agactgaaac ttggagtgtt gttgatagca taagtttaag    1080
51 gtgcattaaa ggattggtac ctggaatctt cccgattgt cagaccggtg agtatgtttt    1140
52 cttggctact cataaacagg ttttggtgta tcaaagacga agtaagttat ggaaagagat    1200
53 gttttctgta aaaggaagct cttctctgcc tttgtggttc tctgctcacg cctttcgcag    1260
54 caccatagta ccctgtaatt agcatgttta tgtttccttc tctactcttt tattttttttg    1320
55 gtttatgttc agctcttgga tcttttaggg cttatgaaaa tttgttcaag gttttataat    1380
56 ctttctggga taacatcata taaagtaatg tacagttgat ttcttctgtt gctttttagta    1440
57 caaatagagt tttggttg 1458
59 <210> SEQ ID NO: 2
```

11/20/01

RAW SEQUENCE LISTING   DATE: 11/20/2001
PATENT APPLICATION: US/09/637,780B   TIME: 09:18:49

Input Set : **N:\EBONY'S\ES.txt**
Output Set: **N:\CRF3\11202001\I637780B.raw**

```
60 <211> LENGTH: 359
61 <212> TYPE: PRT
62 <213> ORGANISM: Arabidopsis thaliana
64 <220> FEATURE:
65 <221> NAME/KEY: misc_feature
66 <222> LOCATION: 1..359
67 <223> OTHER INFORMATION: Xaa is any amino acid
69 <220> FEATURE:
70 <221> NAME/KEY: misc_feature
71 <222> LOCATION: 1..359
72 <223> OTHER INFORMATION: Ceres Seq. ID 1481333
74 <400> SEQUENCE: 2
76 Met Asp Asn Gln Lys Gly Ala Leu Phe Pro Asp Glu Val Ile Leu Gln
77 1               5                  10                  15
78 Ile Leu Ala Arg Leu Pro Val Lys Ser Leu Phe Arg Phe Lys Ser Val
79             20                  25                  30
80 Cys Lys Ser Trp Tyr Arg Leu Pro Ser Asp Lys Tyr Phe Thr Ser Leu
81         35                  40                  45
82 Phe Asn Gln Leu Ser Val Lys Glu Gln Leu Leu Val Ala Gln Val Ser
83     50                  55                  60
84 Asp Ser Ser Ser Leu Ile Cys Val Asp Asn Leu Arg Gly Val Ser Glu
85 65                  70                  75                  80
86 Leu Ser Leu Asp Phe Val Arg Asp Arg Val Arg Ile Arg Val Ser Ser
87             85                  90                  95
88 Asn Gly Leu Leu Cys Cys Ser Ser Ile Pro Glu Lys Gly Val Tyr Tyr
89             100                 105                 110
90 Val Cys Asn Pro Ser Thr Arg Glu Tyr Arg Lys Leu Pro Lys Ser Arg
91         115                 120                 125
92 Glu Arg Pro Val Thr Arg Phe Tyr Pro Asp Gly Glu Ala Thr Leu Val
93     130                 135                 140
94 Gly Leu Ala Cys Asp Leu Ser Arg Asn Lys Phe Asn Val Val Leu Ala
95 145                 150                 155                 160
96 Gly Tyr His Arg Ser Phe Gly Gln Arg Pro Asp Gly Ser Phe Ile Cys
97             165                 170                 175
98 Leu Val Phe Asp Ser Glu Ser Asn Lys Trp Arg Lys Phe Val Ser Val
99             180                 185                 190
100 Leu Glu Glu Cys Ser Phe Thr His Met Ser Lys Asn Gln Val Val Phe
101             195                 200                 205
102 Val Asn Gly Met Leu His Trp Leu Met Ser Gly Leu Cys Tyr Ile Leu
103         210                 215                 220
104 Ala Leu Asp Val Glu His Asp Val Trp Arg Lys Ile Ser Leu Pro Asp
105 225                 230                 235                 240
106 Glu Ile Lys Ile Gly Asn Gly Gly Asn Arg Val Tyr Leu Leu Glu
107                 245                 250                 255
108 Ser Asp Gly Phe Leu Ser Val Ile Gln Leu Ser Asp Val Trp Met Lys
109             260                 265                 270
110 Ile Trp Lys Met Ser Glu Tyr Glu Thr Glu Thr Trp Ser Val Val Asp
111         275                 280                 285
112 Ser Ile Ser Leu Arg Cys Ile Lys Gly Leu Val Pro Gly Ile Phe Pro
```

```
113       290                    295                    300
114 Ile Cys Gln Thr Gly Glu Tyr Val Phe Leu Ala Thr His Lys Gln Val
115 305                    310                    315                    320
116 Leu Val Tyr Gln Arg Arg Ser Lys Leu Trp Lys Glu Met Phe Ser Val
117                    325                    330                    335
118 Lys Gly Ser Ser Ser Leu Pro Leu Trp Phe Ser Ala His Ala Phe Arg
119                    340                    345          ·          350
120 Ser Thr Ile Val Pro Cys Asn
121           355
123 <210> SEQ ID NO: 3
124 <211> LENGTH: 1353
125 <212> TYPE: DNA
126 <213> ORGANISM: Arabidopsis thaliana
128 <220> FEATURE:
129 <221> NAME/KEY: misc_feature
130 <222> LOCATION: 1..1353
131 <223> OTHER INFORMATION: any n = a, g, c, t, unknown, or other
133 <220> FEATURE:
134 <221> NAME/KEY: misc_feature
135 <222> LOCATION: 1..1353
136 <223> OTHER INFORMATION: Ceres Seq. ID 1481342
138 <400> SEQUENCE: 3
140 casragamcc atwacywaga amcaycctaa tcgaaaaaac gccacaatca tggctttgtt      60
141 cttatctcct aaaaccatca ctcttctctt cttctccctc tccctcgcac tctactgcag     120
142 catcgatcct ttccaccact gcgccatttc cgatttcccc aatttcgtct ctcacgaagt     180
143 tatctctcca cgtcccgacg aagttccatg ggagagagat tcacaaaatt cacttcagaa     240
144 atcaaagatt ctgtttttta accaaatcca aggtccagag agcgtcgcct ttgattctct     300
145 cggacgtggt ccgtacacag gcgttgctga tggtagggtt ttgttttggg atggagagaa     360
146 atggattgat ttcgcttata cttcgagtaa tcgatcggag atttgtgatc cgaagccttc     420
147 tgctttgagt tacttgagga atgaacatat atgtggtcgt cctttaggtc ttcgtttcga     480
148 taagagaacc ggagatttgt atatagctga tgcttatatg ggacttttga aagttggtcc     540
149 tgaaggtggt ttagcaacgc cgcttgtaac tgaagctgaa ggtgtgccgt tggggtttac     600
150 taatgatctt gacattgctg atgatggaac tgtttacttt acagatagca gcattagtta     660
151 ccagaggagg aacttcttgc agctcgtttt ctctggagac aatactggga gggttctaaa     720
152 gtatgatcca gtagctaaga aagctgttgt tttggtctca aatcttcagt ttccgaatgg     780
153 tgtctctatc agcagagacg gttctttctt tgtattctgc gaaggagata ttggaagcct     840
154 acgaagatac tggttgaaag gcgagaaagc tggaacgaca gatgtgtttg cgtatttacc     900
155 agggcatcct gataacgtaa gaaccaacca aaagggtgaa ttttgggtag cgcttcattg     960
156 cagacgcaac tactactcat acttaatggc aagatatcct aagctgagga tgttcatact    1020
157 gagactgcca atcactgcga gaactcacta ctcgttccag ataggttac ggccgcacgg    1080
158 gttggtggtt aagtatagtc ctgaagggaa gcttatgcat gttttggaag atagtgaagg    1140
159 gaaagttgtg agatcagtaa gtgaagtgga agaaaaagat gggaagcttt ggatgggaag    1200
160 tgtgttgatg aactttgttg ctgtctatga cctctgatta cttgacctat acgtaaacca    1260
161 cttcactcag tttctagatt tagcaaattc tcaaaactgt taggtgtgta ctgaaaaaat    1320
162 caaacactta gcacaaacaa actcaatgtt att 1353
164 <210> SEQ ID NO: 4
165 <211> LENGTH: 411
166 <212> TYPE: PRT
167 <213> ORGANISM: Arabidopsis thaliana
```

RAW SEQUENCE LISTING   DATE: 11/20/2001
PATENT APPLICATION: US/09/637,780B   TIME: 09:18:49

Input Set : N:\EBONY'S\ES.txt
Output Set: N:\CRF3\11202001\I637780B.raw

```
169 <220> FEATURE:
170 <221> NAME/KEY: misc_feature
171 <222> LOCATION: 1..411
172 <223> OTHER INFORMATION: Xaa is any amino acid
174 <220> FEATURE:
175 <221> NAME/KEY: misc_feature
176 <222> LOCATION: 1..411
177 <223> OTHER INFORMATION: Ceres Seq. ID 1481343
179 <400> SEQUENCE: 4
```

W--> 181 Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Pro Asn Arg Lys Asn Ala Thr Ile
     182 1   5   10   15
     183 Met Ala Leu Phe Leu Ser Pro Lys Thr Ile Thr Leu Leu Phe Phe Ser
     184   20   25   30
     185 Leu Ser Leu Ala Leu Tyr Cys Ser Ile Asp Pro Phe His His Cys Ala
     186   35   40   45
     187 Ile Ser Asp Phe Pro Asn Phe Val Ser His Glu Val Ile Ser Pro Arg
     188   50   55   60
     189 Pro Asp Glu Val Pro Trp Glu Arg Asp Ser Gln Asn Ser Leu Gln Lys
     190 65   70   75   80
     191 Ser Lys Ile Leu Phe Phe Asn Gln Ile Gln Gly Pro Glu Ser Val Ala
     192   85   90   95
     193 Phe Asp Ser Leu Gly Arg Gly Pro Tyr Thr Gly Val Ala Asp Gly Arg
     194   100   105   110
     195 Val Leu Phe Trp Asp Gly Glu Lys Trp Ile Asp Phe Ala Tyr Thr Ser
     196   115   120   125
     197 Ser Asn Arg Ser Glu Ile Cys Asp Pro Lys Pro Ser Ala Leu Ser Tyr
     198   130   135   140
     199 Leu Arg Asn Glu His Ile Cys Gly Arg Pro Leu Gly Leu Arg Phe Asp
     200 145   150   155   160
     201 Lys Arg Thr Gly Asp Leu Tyr Ile Ala Asp Ala Tyr Met Gly Leu Leu
     202   165   170   175
     203 Lys Val Gly Pro Glu Gly Gly Leu Ala Thr Pro Leu Val Thr Glu Ala
     204   180   185   190
     205 Glu Gly Val Pro Leu Gly Phe Thr Asn Asp Leu Asp Ile Ala Asp Asp
     206   195   200   205
     207 Gly Thr Val Tyr Phe Thr Asp Ser Ser Ile Ser Tyr Gln Arg Arg Asn
     208   210   215   220
     209 Phe Leu Gln Leu Val Phe Ser Gly Asp Asn Thr Gly Arg Val Leu Lys
     210 225   230   235   240
     211 Tyr Asp Pro Val Ala Lys Lys Ala Val Val Leu Val Ser Asn Leu Gln
     212   245   250   255
     213 Phe Pro Asn Gly Val Ser Ile Ser Arg Asp Gly Ser Phe Phe Val Phe
     214   260   265   270
     215 Cys Glu Gly Asp Ile Gly Ser Leu Arg Arg Tyr Trp Leu Lys Gly Glu
     216   275   280   285
     217 Lys Ala Gly Thr Thr Asp Val Phe Ala Tyr Leu Pro Gly His Pro Asp
     218   290   295   300
     219 Asn Val Arg Thr Asn Gln Lys Gly Glu Phe Trp Val Ala Leu His Cys
     220 305   310   315   320

RAW SEQUENCE LISTING   DATE: 11/20/2001
PATENT APPLICATION: US/09/637,780B   TIME: 09:18:49

Input Set : N:\EBONY'S\ES.txt
Output Set: N:\CRF3\11202001\I637780B.raw

```
221 Arg Arg Asn Tyr Tyr Ser Tyr Leu Met Ala Arg Tyr Pro Lys Leu Arg
222                 325             330             335
223 Met Phe Ile Leu Arg Leu Pro Ile Thr Ala Arg Thr His Tyr Ser Phe
224             340             345             350
225 Gln Ile Gly Leu Arg Pro His Gly Leu Val Val Lys Tyr Ser Pro Glu
226         355         .       360             365
227 Gly Lys Leu Met His Val Leu Glu Asp Ser Glu Gly Lys Val Val Arg
228     370             375             380
229 Ser Val Ser Glu Val Glu Glu Lys Asp Gly Lys Leu Trp Met Gly Ser
230 385             390             395                 400
231 Val Leu Met Asn Phe Val Ala Val Tyr Asp Leu
232             405             410
234 <210> SEQ ID NO: 5
235 <211> LENGTH: 395
236 <212> TYPE: PRT
237 <213> ORGANISM: Arabidopsis thaliana
239 <220> FEATURE:
240 <221> NAME/KEY: misc_feature
241 <222> LOCATION: 1..395
242 <223> OTHER INFORMATION: Xaa is any amino acid
244 <220> FEATURE:
245 <221> NAME/KEY: misc_feature
246 <222> LOCATION: 1..395
247 <223> OTHER INFORMATION: Ceres Seq. ID 1481344
249 <400> SEQUENCE: 5
251 Met Ala Leu Phe Leu Ser Pro Lys Thr Ile Thr Leu Leu Phe Phe Ser
252 1               5               10              15
253 Leu Ser Leu Ala Leu Tyr Cys Ser Ile Asp Pro Phe His His Cys Ala
254             20              25              30
255 Ile Ser Asp Phe Pro Asn Phe Val Ser His Glu Val Ile Ser Pro Arg
256         35              40              45
257 Pro Asp Glu Val Pro Trp Glu Arg Asp Ser Gln Asn Ser Leu Gln Lys
258     50              55              60
259 Ser Lys Ile Leu Phe Phe Asn Gln Ile Gln Gly Pro Glu Ser Val Ala
260 65              70              75              80
261 Phe Asp Ser Leu Gly Arg Gly Pro Tyr Thr Gly Val Ala Asp Gly Arg
262             85              90              95
263 Val Leu Phe Trp Asp Gly Glu Lys Trp Ile Asp Phe Ala Tyr Thr Ser
264             100             105             110
265 Ser Asn Arg Ser Glu Ile Cys Asp Pro Lys Pro Ser Ala Leu Ser Tyr
266         115             120             125
267 Leu Arg Asn Glu His Ile Cys Gly Arg Pro Leu Gly Leu Arg Phe Asp
268     130             135             140
269 Lys Arg Thr Gly Asp Leu Tyr Ile Ala Asp Ala Tyr Met Gly Leu Leu
270 145             150             155             160
271 Lys Val Gly Pro Glu Gly Gly Leu Ala Thr Pro Leu Val Thr Glu Ala
272             165             170             175
273 Glu Gly Val Pro Leu Gly Phe Thr Asn Asp Leu Asp Ile Ala Asp Asp
274             180             185             190
```

Use of n and / or Xaa has been detected in the
Sequence Listing. Review the Sequence Listing
to ensure a corresponding explanation is present
in the <220> to <223> fields of each sequence
using n or Xaa.

VERIFICATION SUMMARY   DATE: 11/20/2001
PATENT APPLICATION: **US/09/637,780B**   TIME: 09:18:50

Input Set : **N:\EBONY'S\ES.txt**
Output Set: **N:\CRF3\11202001\I637780B.raw**

```
L:181  M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:1112 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26
L:1146 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27
L:1180 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28
L:1210 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:29
L:1594 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:39
L:1634 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:40
L:1674 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41
L:4873 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:108
L:4929 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:109
L:4977 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:110
L:5200 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:115
L:5202 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:115
L:5232 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:116
L:5236 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:116
L:5299 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:118
L:5558 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:125
L:5584 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:126
L:5596 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:126
L:5620 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:127
L:5624 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:127
L:5626 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:127
L:5656 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:128
L:5960 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:138
L:5966 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:138
L:5968 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:138
L:6002 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:139
L:6008 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:139
L:6010 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:139
L:6044 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:140
L:6048 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:140
L:6050 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:140
L:6052 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:140
L:6202 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:145
L:6242 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:146
L:6280 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:147
L:6309 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:148
L:6329 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:149
L:6331 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:149
L:6333 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:149
L:6335 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:149
L:6384 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:151
L:6394 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:151
L:6418 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:152
L:6428 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:152
L:6487 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:154
L:6489 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:154
L:6491 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:154
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/09/637,780B**

DATE: 11/20/2001
TIME: 09:18:50

Input Set : **N:\EBONY'S\ES.txt**
Output Set: **N:\CRF3\11202001\I637780B.raw**

L:6493 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:154
L:6495 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:154